



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2021

Integrating behavioral and geospatial data on the timeline: towards new dimensions of analysis

Larsson, Jens ; Burenhult, Niclas ; Kruspe, Nicole ; Purves, Ross S ; Rothstein, Mikael ; Sercombe, Peter

Abstract: Studies of human spatial behavior increasingly rely on a combination of audiovisual and geospatial recordings. So far, however, few analytical environments have offered opportunities for integrated and expedient annotation and analysis of the two. Here we report the first study aimed at integrating geospatial data in an environment developed for time-aligned annotation of audiovisual media. By calibrating the audiovisual and geospatial signals on the timeline and inserting the geo data as a tier in the annotation tool ELAN, we innovatively generate an environment in which time-aligned annotations of audiovisually observed behavior can be linked and explored in relation to the corresponding geographical coordinates. We illustrate the technique with cultural and linguistic behavior recorded on the move among indigenous communities in Southeast Asia. Our methodological principle is of potential interest to any study or discipline concerned with linking the location and properties of observable behavior.

DOI: <https://doi.org/10.1080/13645579.2020.1763705>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-193189>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Larsson, Jens; Burenhult, Niclas; Kruspe, Nicole; Purves, Ross S; Rothstein, Mikael; Sercombe, Peter (2021). Integrating behavioral and geospatial data on the timeline: towards new dimensions of analysis. *International Journal of Social Research Methodology*, 24(1):1-13.

DOI: <https://doi.org/10.1080/13645579.2020.1763705>



Integrating behavioral and geospatial data on the timeline: towards new dimensions of analysis

Jens Larsson, Niclas Burenhult, Nicole Kruspe, Ross S. Purves, Mikael Rothstein & Peter Sercombe

To cite this article: Jens Larsson, Niclas Burenhult, Nicole Kruspe, Ross S. Purves, Mikael Rothstein & Peter Sercombe (2020): Integrating behavioral and geospatial data on the timeline: towards new dimensions of analysis, International Journal of Social Research Methodology, DOI: [10.1080/13645579.2020.1763705](https://doi.org/10.1080/13645579.2020.1763705)

To link to this article: <https://doi.org/10.1080/13645579.2020.1763705>



© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 15 May 2020.



Submit your article to this journal [↗](#)



Article views: 461



View related articles [↗](#)



View Crossmark data [↗](#)

Integrating behavioral and geospatial data on the timeline: towards new dimensions of analysis

Jens Larsson^a, Niclas Burenhult^{a,b}, Nicole Kruspe^b, Ross S. Purves^c, Mikael Rothstein^d and Peter Sercombe^e

^aLund University Humanities Lab, Lund University, Lund, Sweden; ^bCentre for Languages and Literature, Lund University, Lund, Sweden; ^cDepartment of Geography/University Research Priority Program for Language and Space, University of Zurich, Zurich, Switzerland; ^dSection for the Study of Religions/Department of History, University of Southern Denmark, Odense, Denmark; ^eSchool of Education, Communication and Language Sciences, Newcastle University, Newcastle, UK

ABSTRACT

Studies of human spatial behavior increasingly rely on a combination of audiovisual and geospatial recordings. So far, however, few analytical environments have offered opportunities for integrated and expedient annotation and analysis of the two. Here we report the first study aimed at integrating geospatial data in an environment developed for time-aligned annotation of audiovisual media. By calibrating the audiovisual and geospatial signals on the timeline and inserting the geo data as a tier in the annotation tool ELAN, we innovatively generate an environment in which time-aligned annotations of audiovisually observed behavior can be linked and explored in relation to the corresponding geographical coordinates. We illustrate the technique with cultural and linguistic behavior recorded on the move among indigenous communities in Southeast Asia. Our methodological principle is of potential interest to any study or discipline concerned with linking the location and properties of observable behavior.

ARTICLE HISTORY

Received 24 January 2020
Accepted 29 April 2020

KEYWORDS

Spatial behavior;
georeferencing; audiovisual
media; annotation; ELAN

1. Introduction

Audiovisual recording is a well-established method in the humanities and social sciences (Ashmore, 2008; Bates, 2015; Jewitt, 2012; Margolis & Pauwels, 2011; Spencer, 2010). Likewise, geospatial recording with Global Positioning System (GPS) receivers is becoming an increasingly utilized method for capturing and analyzing a range of expressions of human culture and behavior (Abernathy, 2016; Gawne & Ring, 2016; Spencer et al., 2003). Thus far, however, these two dimensions of recording have largely represented separate endeavors. This is partly due to technological limitations: until recently, a lack of recording devices which capture both the audiovisual and geospatial signal made parallel recording impractical. Furthermore, the stationary procedures and set-ups of audiovisual recording did not generate an obvious conceptual impetus for simultaneous recording of location. With the recent advent of multi-format mobile recording devices, such as action cameras, this situation has changed. Today, compact devices, designed to record recreational action on-the-go, can capture high-quality video and audio and continuous geodata in tandem, clearing the path for new applications in research. Usage for scientific purposes in the humanities

and social sciences has been limited so far but this is likely to change in the near future (see, e.g., Cialone, 2019).

An obstacle to this development is the lack of dedicated analytical environments in which synchronous audiovisual and geospatial data can be integrated and explored. Our aim in this contribution is to present a cost-effective method and protocol for time-aligned annotations of such data. The proposed workflow is based on mobile and dynamic environments where audiovisual and geospatial data are captured in parallel. In order to map audiovisual events that occur over both time and space to dynamic geographical locations, we need an interface and a workflow that can host and integrate the two.

Continuously recorded geospatial data are point-based and are effectively a temporal data stream: each logged point is timestamped and can be represented as plain text on a timeline. This is why integration with existing tools for time-aligned, linguistic annotations are of special interest, since this ensures that users do not have to relearn software-specific workflows.

Our software of choice is ELAN.¹ Developed by The Language Archive at the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands, this is an advanced tool for annotating events in audiovisual data, exactly when they occur, for the duration of the event (Sloetjes & Seibert, 2016; Wittenburg et al., 2006). It is free and open-source, while its file format (EAF)² is an open XML³ specification that can be parsed, generated and validated reliably. Members of our team also have previous experience with ELAN, making it a prime candidate for this pilot (see illustration of ELAN in Figure 1).

In the following sections, we describe the technological and procedural aspects of recording, and the extraction of the geospatial data and their integration in the annotation environment (Section 2). We then illustrate the diverse potential of the technique with linguistic and cultural data recorded on the move among indigenous communities in Borneo and the Malay Peninsula

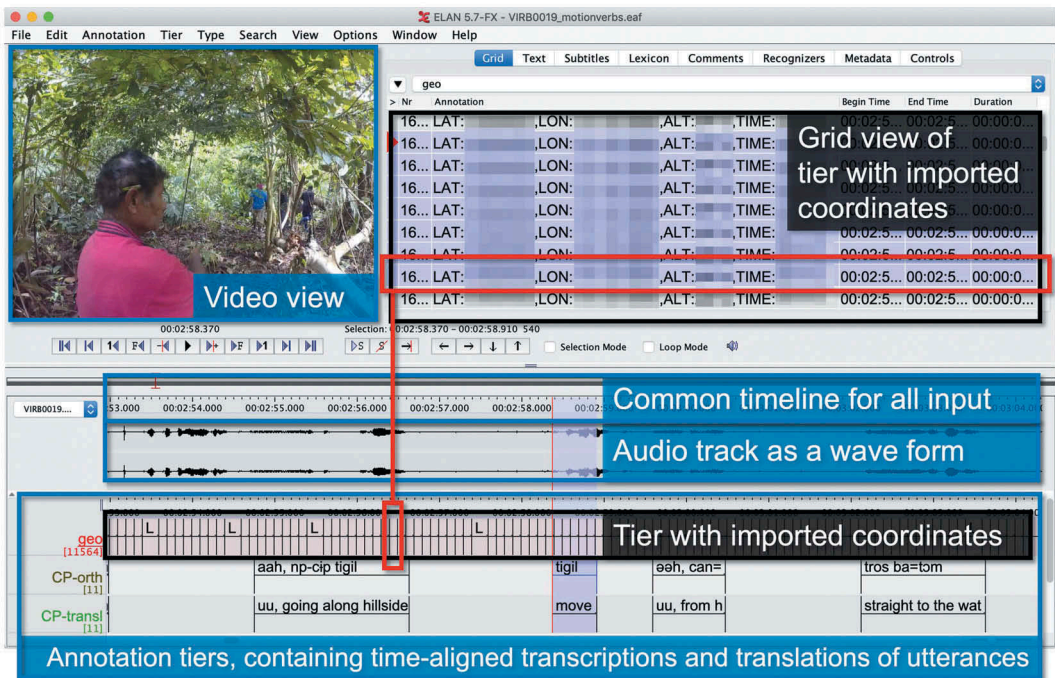


Figure 1. An ELAN view, illustrating the integration of audio, video, multi-tiered linguistic annotations and geographical coordinates on a common timeline. Standard ELAN functionalities are highlighted in blue, our own georeferential implementations in black. The coordinate values have been obscured for privacy.

(Section 3). We conclude with a discussion of possible applications, an assessment of the conceptual and theoretical rewards as well as the ethical ramifications that our approach entails (Section 4).

Our approach represents a conceptually significant leap in that it combines two well established but hitherto analytically unconnected recording formats in an annotation environment. As we will show, its principles pave the way for a new line of high-resolution analysis in the spatio-temporal dimensions, of potential interest to any study or discipline concerned with linking the location and properties of observable behavior.

2. Method

2.1. Recording devices and techniques

For this pilot, our recording device is an action camera, strapped to one's chest. Designed for activities such as climbing or skiing, these cameras are not only rugged and easy to operate but also allow for the continuous logging of position via a built-in GPS. We have opted for the Garmin VIRB Ultra 30 since it utilizes a well-established, multipurpose format for telemetry, capable of logging anything from heart rate to geospatial data on a common timeline. Moreover, the VIRB logs telemetry in the background, whether a video is being recorded or not, but still keeps everything in sync. Together with the fact that documentation and developer tools for the data format are also freely available, the VIRB becomes a data hub well suited for our pilot.

2.2. Principles and procedures of geospatial data extraction, insertion and integration

While simultaneously capturing video and geospatial data is a one-button affair with the current breed of action cameras, some form of data synchronization is usually necessary. For a video clip lasting 10 min, we have to identify which part of the geospatial data stream those 10 min correspond to. This requires a common timeline.

GPS logs commonly contain both date and time for each logged point, whereas digital video cameras often embed this information directly inside the video files. Our annotation tool ELAN, on the other hand, only has a relative timeline. Annotations are tied to time stamps that denote boundaries for when an annotated event begins and ends in relation to the beginning of the media stream. Consequently, inside ELAN we cannot easily determine what date and time 2 minutes, 43 seconds into a recording corresponds to. To overcome this, our method takes full advantage of the specialized telemetry format utilized by the VIRB, allowing for higher temporal precision and a fully automated synchronization process.⁴

The telemetry data are stored as a separate file in the so-called FIT format,⁵ which generates a common timeline where all logged events, such as logged coordinates, are individually timestamped. In order to parse and export data logged in this format, post-processing is required. For synchronization purposes, the VIRB embeds a unique identifier into every video file. In the FIT file, this identifier is logged in time-stamped data messages denoting the start and end of the corresponding clip. This allows us to identify any logged data message that intersects with the time span of each, respective video clip, and determine exactly when in the video this message was logged. We have developed simple proof-of-concept tools to automate the process that parses the raw telemetry data directly for more control over what we extract and the structure of the exported data. ELAN's relative time line can now be shifted to the absolute time line we established for the video and its corresponding set of geospatial data.

Currently, ELAN has no ideal option for importing geospatial data. Therefore, an annotation tier is our best candidate as a data insertion point for this pilot. Each logged coordinate corresponds to one annotation containing latitude, longitude, altitude and time represented as text. On this tier, the start time of each respective annotation represents the time for when a point was originally logged in relation to the video's time line (Figure 1). Their durations correspond to the pauses between samples, e.g., for a 10 Hz GPS sample rate, each annotation is expected to have a duration of around

100 ms. In cases where the GPS for some reason has not been able to log a point, the annotation's duration will simply extend to the time of the next logged point.

Any annotation on a parallel tier can now be georeferenced by determining how many logged points its time frame intersects with. That is, we can georeference events observed in the audiovisual data simply by annotating them. These events can represent a single location or a path.

Note that for longer recordings, one may want to avoid directly importing the full GPS log into ELAN and instead georeference annotations by synchronizing directly against the geospatial data in the FIT file. On the other hand, the ELAN annotation format can perhaps be seen as a test bed for integrating and annotating temporal data types of low and medium resolution.

2.3. Selection, extraction, export and analysis of geospatial data from the EAF-file

To be able to georeference events that are of both single-point and multi-point/line nature, we need an output format that covers these cases. For this pilot, KML⁶ is our choice since it is both human and computer readable, it is well documented, it can be validated and is compatible with various kinds of Geographic Information Systems (GIS) software.

This leads us to a few design decisions that affect granularity and precision of the output. In ELAN, millisecond precision is often crucial (e.g., when annotating a gesture, or when aligning a spoken word with a transcription). However, the granularity of the geospatial data will impose limitations in terms of how accurately an observed event can be represented in time and space. If the GPS device logs one point per second, this also becomes the limit for temporal precision. For spatial precision, there are the limitations of the GPS technology itself to consider. At lower speeds, for longer time spans or events that represent static concepts, such as place names, this might not be an issue, however. The VIRB logs and time stamps 10 points per second which does generate fairly large amounts of data, but in return, we gain slightly finer granularity.

3. Case studies

In this section, we illustrate how the integration of aligned audiovisual recordings, geospatial data and annotations can be put to use with examples of data sets from three indigenous communities in Southeast Asia: the Penan of Borneo, and the Jahai and Semelai of the Malay Peninsula. All three cases involve on-the-move recordings with chest-mounted action cameras; however, they illustrate distinct types of analytical targets, capturing some of the wide range of potential applications.

3.1. Eastern Penan: knowledge and experience of place

Our first example comes from ongoing work by co-authors Rothstein and Sercombe on knowledge and experience of place among the Penan, a group of a few thousand hunter-gatherers in Brunei and Sarawak, North-Central Borneo, who speak an Austronesian language. Only a very small number of Penan still live as nomadic foragers (Rothstein, 2016; Sercombe & Sellato, 2007). Traditionally they live in egalitarian groups of 20–40 individuals, travelling through the group's customary land, *tana' pengurip*, an area which would typically cover upwards of 100 square kilometers.

The Penan traditional lands are undergoing rapid development, especially deforestation, and for the majority of the Penan the environmental changes, coupled with intense efforts to modernize and Christianize them, mean that traditional customs and knowledge are quickly disappearing. The present research program aims to capture and document some of the traditional knowledge relating to *tana' pengurip*, and our example represents a pilot study of in situ environmental descriptions and personal narratives by Penan custodians of this knowledge.

The data were obtained from a part of the Penan territory located in the Kelabit Highlands. Two men, each carrying a camera, were asked to follow a trajectory of their own choice together. They were asked to relate to the landscape as they went along, and, when relevant, exchange knowledge

and experiences in their own language pertaining to the places they would pass. The walk took the Penan to areas that were well known to them but, given their recent sedentary condition, not part of their immediate surroundings.

Figure 2 illustrates an excerpt where the two men traverse a plateau in an area known as Pawan, following a traditional Penan path that they know well and recounting matters of landscape significance to the Penan. At one point a rocky, tree-clad eminence is about to come into the view of the men and the cameras, which triggers immediate commentary by one of the men. He mentions its Kelabit name, Pawan, and goes on to explain that in Penan it is referred to as *Gerusu* 'Rocky'. After a few minutes of silent motion towards the feature, and upon reaching it, he starts to recount traditional ways associated with the site and the area, first commenting on the site's significance as a frequent dwelling place when the Penan were still nomadic. He then explains what local camps and households were like. After a short period of silence, he describes the traditional way of calling out to people at a distance for the purpose of meeting up.

Following transcription and translation in ELAN, and the insertion of the geodata as an annotation tier (see Section 2.2), we are in a position to make a precise analysis of the relationship between the viewpoints, the different instances of commentary, and the geographical coordinates of those viewpoints and commentaries. For example, the location of the commentary triggered by the initial approach to the site is georeferenced at a distance of approximately 200 m from the feature referred to. Subsequent commentary is clustered at or very near the same feature. Thus, the narrative can be explored as it unfolds both temporally and spatially, and its different components triangulate on a place deemed significant by the participants. The relationship between commentary and location is illustrated in Figure 2.

Any assessment of the social, cosmological and material reality of the nomadic Penan must take movement – the nomadic process – into account. The integration of audiovisual, geospatial and

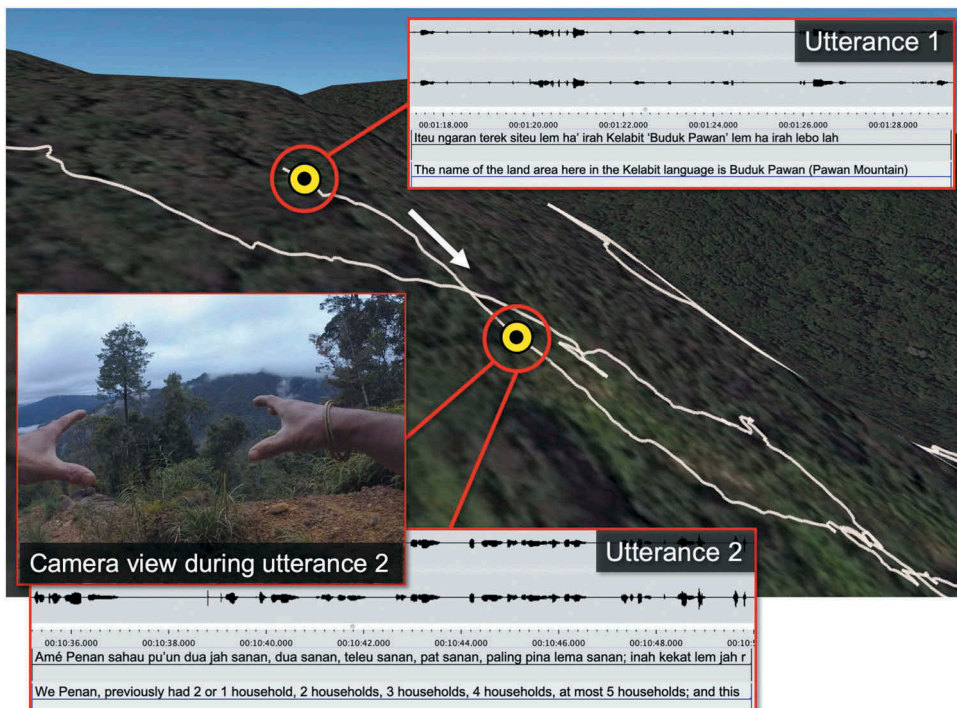


Figure 2. Georeferenced view and ELAN annotations of utterances from *in situ* Penan spatial narratives. The corresponding coordinates are illustrated as points in Google Earth. (Google, Maxar Technologies, Landsat/Copernicus).

linguistic data illustrated here enables us to do that and to create an unparalleled environment for exploration and documentation of the Penan's relationship with their surroundings.

3.2. *Jahai: spatial categories in language*

Our second example comes from ongoing work by co-author Burenhult on spatial categories in the language of the Jahai, subsistence foragers in the rainforests of the Malay Peninsula speaking an Austroasiatic language. Two linguistic categories are targeted: indigenous place names and verbs denoting motion in relation to landscape features. The two types both involve focused extraction of instances of the categories in question. However, they differ in that place names represent actual and stationary referents along a route, whereas the motion verbs represent sections of the route and encode the movements made. For this reason, different elicitation and recording approaches were chosen when capturing the two.

3.2.1. *Place names*

The collection of Jahai place names forms part of a wider effort to document such names and to understand the principles of naming. Action cameras were employed to make it possible for community members to engage in on-site geospatial recording of names while on the move, without having to operate a separate GPS unit. At the same time, they recorded their conversations about the places in question audiovisually, again without having to actively operate a video camera (since the cameras kept recording throughout the journey).

The method was piloted by a foraging party moving on foot through a watershed in the Mendelum valley, in Perak, Peninsular Malaysia. Two participants carried synchronously recording cameras and received prior instructions from the researcher (who did not participate in the trip) to mention the name of each named place they passed through, and to do so while they were at that specific place. Jahai place names lend themselves well to this type of collection, especially if the trip occurs along a major stream. This is because names correspond to watersheds of such streams and their tributaries, and lengthwise motion along a stream makes it possible to exhaustively collect the names of tributary areas. The recorded part of the trip lasted about 1.5 hours, ending when the camera batteries discharged.

The recorded conversations were subsequently transcribed and translated in ELAN, and the geodata were added as a separate tier. The ensuing analysis of the transcriptions and audiovisual data identified 15 instances of place names uttered in situ. These were isolated and selected, and the corresponding coordinates on the timeline were exported to a KML file. [Figure 3](#) illustrates the relationship between transcribed place names and the corresponding coordinates in ELAN, and the extracted names in Google Earth.

The example highlights the possibility to identify and locate stationary spatial referents through combined audiovisual, geospatial and annotated information coinciding on the timeline. Although the geospatial points or sequences selected do not match the full extent of the named referents (whose spatial limits are vague and challenging to capture in detail), they pinpoint the in situ uttering of the relevant linguistic exemplar. Our example is place names, but other types of spatial referents are equally recordable with this method, such as landforms or ecotopes.

3.2.2. *Motion categories*

Our second Jahai example illustrates motion events. The Jahai language has a set of verbs which express locomotion in relation to geographical features (Burenhult & Purves, 2020). Each motion verb associates with a particular type of feature and also expresses the path of movement in relation to that feature (up, down, along, across), e.g., *tigil* 'move-across-hillside', *rkruk* 'move-along-major-stream'. The goal is to create an environment in which real instantiations of the verbs can be analyzed in relation to local topography and hydrology.

Instantiations of the verbs are recordable with a GPS as lines, which trace the particular type of movement involved. However, although representing concrete spatial manifestations, the

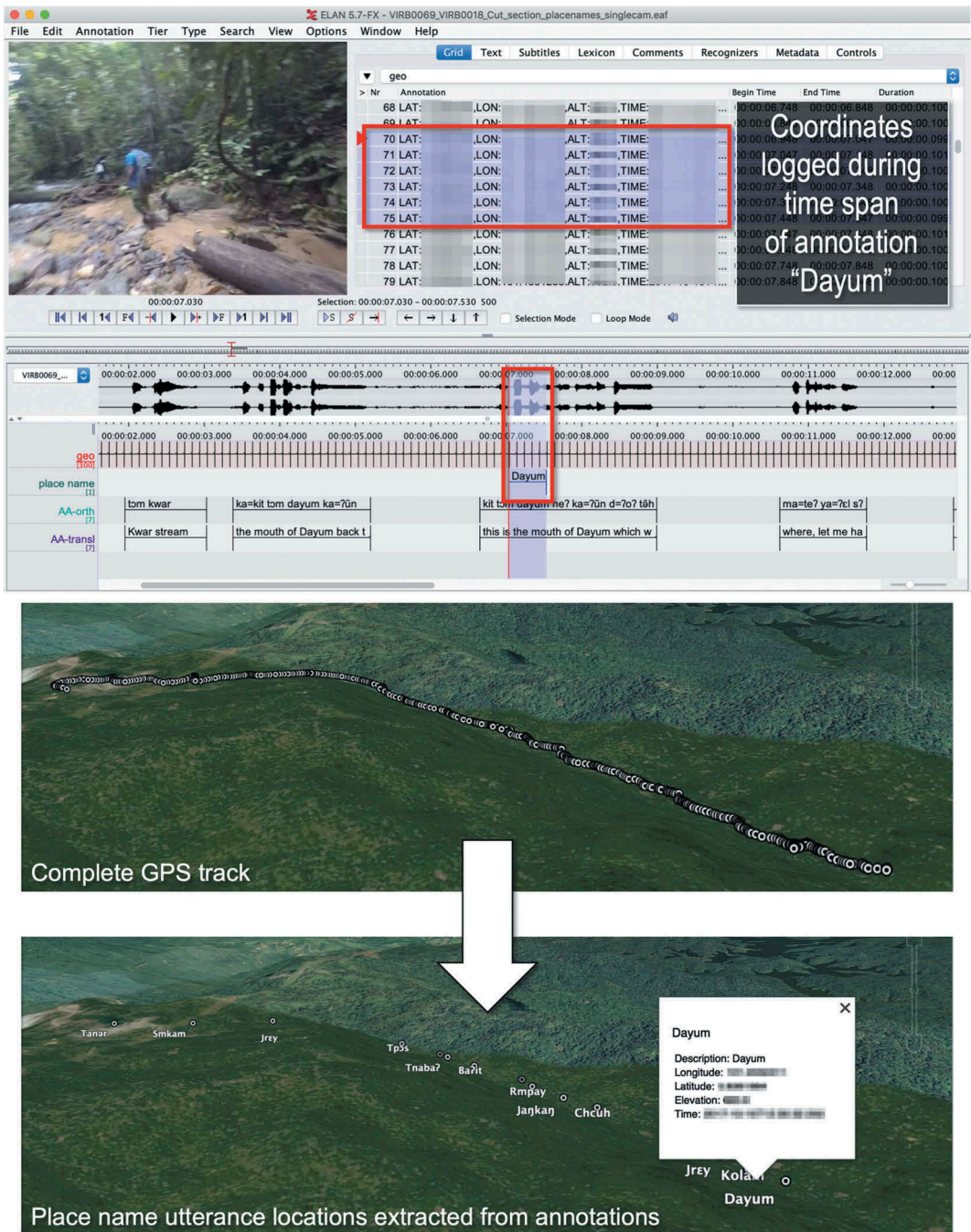


Figure 3. Top panel: Selection and georeferencing of an ELAN annotation of a Jahai place name uttered *in situ*. Bottom panel: The coordinates corresponding to such annotations extracted from the logged track and illustrated as points in Google Earth. The two steps illustrate the workflow involved in linguistic annotation and subsequent extraction of corresponding coordinates. The coordinate values have been obscured for privacy. (Google, Landsat/Copernicus, Data SIO, NOAA, U.S. Navy, NGA, GEBCO).

movements are ephemeral and, unlike place names, referentially less obvious for native speaker consultants to specify systematically in conversation. For this reason, a more structured approach was necessary in order to make the data collection exhaustive: the researcher joined a foraging party

moving through the watershed of Raba, Perak, himself wearing an action camera. Throughout the walk, he continually interviewed one of the Jahai participants about which kind of motion was taking place, keeping track of each linguistically labeled event as it started, unfolded, and ended.

We then transcribed and translated the conversation between the researcher and the Jahai consultant in ELAN, and added a separate tier with annotations of the full temporal extent of each exemplar of motion, i.e. the events themselves rather than the linguistic utterances describing them, labeled with the corresponding indigenous verbs. Following the addition of the geodata tier, the coordinate sequences of the motion exemplars were selected and exported to a KML file. Figure 4 illustrates the relationship between annotated motion exemplars and the corresponding coordinates in ELAN.

This example illustrates how the technique can be used to combine spoken information in recordings with observed associated behavior as it unfolds in time and space. This data integration is significant because it makes possible a detailed mapping and spatial analysis of the non-linguistic physical but ephemeral manifestations of abstract event types. Types of motion may be the most obvious phenomenon to capture this way, but one can imagine a range of other targets, e.g., techniques of hunting and wayfinding.

3.3. Semelai: ephemeral percepts in language and ritual

Our final example illustrates the capturing of ephemeral percepts and participant responses to them. It is taken from an ongoing comparative project by co-author Kruspe that documents the language people employ to talk about their sonic environments. The data is drawn from the Semelai, an Austroasiatic-speaking group of the Malay Peninsula. The Semelai traditionally practised swidden horticulture in secondary forest along the waterways of Tasek Bera – a once vast peat swamp forest ecosystem in the state of Pahang – and supplemented this with trade in forest products collected from the surrounding primary forest that had dominated the area.



Figure 4. Selection and georeferencing of an ELAN annotation of a continuous Jahai motion event. The corresponding coordinates are illustrated as a section of a track in Google Earth. The coordinate values have been obscured for privacy. (Google, CNES/Airbus).

The Semelai are animists and share their world with an array of potentially dangerous super-human beings. Both people and superhumans constantly monitor each other's presence in the environment, and this impacts upon the way in which humans conduct themselves, especially in the forest where the beings are particularly prevalent. Humans avoid disturbing them by adhering to prescribed rituals. The researcher's existing documentary corpus contained accounts of events involving sonic perception and descriptions of the linguistic routines employed; the specific aim here was to capture naturalistic in situ examples of 'perception events' in the surrounding environment in real time. Two consultants were fitted with cameras to record them undertaking daily activities outside of the village environment. No specific requests were made to structure the content of the recordings, but consultants were asked to provide place names as they moved through them. The researcher fitted the cameras on the participants and provided instructions on their operation prior to departure, but was never present during the recordings.

In the example given here (Figure 5), two male consultants wore the cameras on a daytime angling trip. They traveled by motorboat downstream on the lake, stopping to fish from the boat in a series of locations. The resulting recordings were then transcribed in ELAN, and the geospatial data were entered as a separate tier. In addition, separate tiers were created to annotate the soundscape, and the uttered place names.

The ambient sounds of the lake are clearly audible in the recording: fish leaping in the water, insect and bird calls, the rustle of the breeze through the tall stands of pandanus (*Pandanus helicopus*), the engines of other boats. At one point, the highly salient two note call of a male great argus pheasant (*Argusianus argus*; Semelai *kawwɔŋ*) is heard. The call of this bird, which can augur ill, is readily identifiable by speakers. It features in Semelai ethnohistorical narratives: the Semelai culture-hero *Ga buŋsu?* gave birth to seven *kawwɔŋ* children when married to her avian

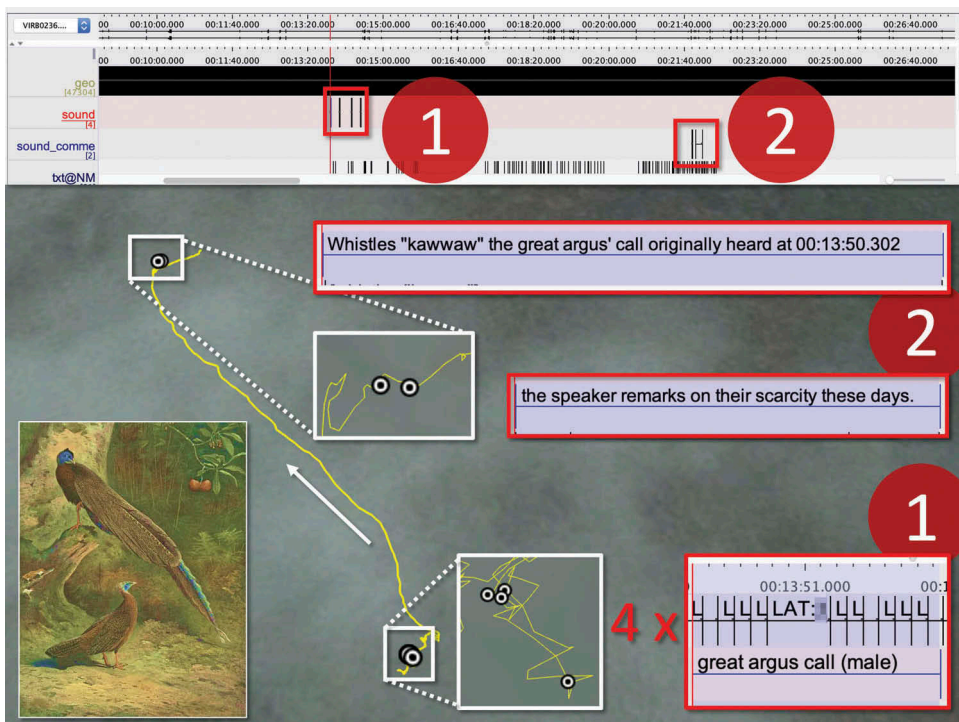


Figure 5. Selection and georeferencing of ELAN annotations of bird calls and subsequent commentary by Semelai participants. The coordinates corresponding to the calls and commentary illustrated as points in Google Earth. The coordinate values have been obscured for privacy. (Google, CNES/Airbus; Painting from The New York Public Library, <https://digitalcollections.nypl.org/items/510d47d9-6ede-a3d9-e040-e00a18064a99>).

husband. She placed each child on a hill in the Semelai landscape. This bird is also represented in shamanic healing rituals, naming the junction to a side path in the underworld that leads to death (Gianno, 2016).

In the recording, the great argus calls suddenly four times over the space of 40 s, then falls silent again. Only after the fourth call does the elder speaker acknowledge the sound. Here we directly observe the speaker enacting a much-reported prohibition: one should not acknowledge a sound from a source which is not visually accessible until it is heard at least three times. Shortly after he briefly muses to his companion that it sounds like the pheasant from a known location back upstream. In a new location, 8 min after hearing the initial call, the elder speaker suddenly whistles, mimicking the great argus's call. He confirms that what they heard was the great argus, then remarks that it is the last one of its kind on the lake. The combined data in ELAN represent a fine-grained analysis of the spatial and temporal relationship between the four bird calls, the participants' hearing them and subsequent commentary on them (Figure 5).

The example illustrates how mobile recording offers an intriguing arena for exploring in situ, instantaneous human reaction to, and representation of transient phenomena experienced while in motion. It is a genre which generates on-the-spot commentary and conversation unlikely to occur in traditional static recording contexts. It should be particularly appropriate for capturing indigenous knowledge of life forms encountered only in the wild. The simultaneous geospatial recording, integrated in the annotation environment, allows for fine-grained documentation and analysis of how the sensory experiences and representations unfold in space and time.

4. Discussion

In this paper, we have described the development of a paradigm for the integration and exploration of synchronous audiovisual and geospatial data. Fundamental to this is the temporal baseline, a dimension of reference along which any continuously recorded data can potentially be structured. A time-based tool for data integration like ELAN is highly suited for this purpose and we here pioneer the linking of continuously recorded geographical coordinates along the timeline in such a tool. Crucial to the analytical usefulness of ELAN is the possibility to create complex annotations on multiple tiers along the one timeline. Any phenomenon perceivable in the audiovisual recording can be represented as a time-aligned annotation on a designated annotation tier. Annotated sequences or categories of analytical interest are selected and matched to the geographical coordinates for precise spatial identification and analysis.

We have illustrated the paradigm with pilot case studies involving intangible cultural and linguistic phenomena of spatial relevance in three indigenous communities. Our examples of georeferenced phenomena are diverse and include place-induced narratives, in situ utterances of static place referents, spatially and temporally dynamic ephemeral actions, and perceived environmental events associated with culturally informative commentary. The annotated phenomena range from millisecond spatial snapshots of single words or utterances to temporally extended narratives or conversations which may or may not progress spatially. Annotated observations may be auditory, visual, or a combination of the two.

ELAN was developed as a tool for linguistic annotation, and our examples point primarily to a broad range of applications involving human language behavior in space. However, as the examples of motion events and bird calls make clear, any phenomenon observed in the audiovisual recordings can be annotated and explored geospatially. This opens up possibilities for applications beyond the language and cultural sciences. Furthermore, our examples are principally aimed to show how behavioral phenomena can be spatialized, answering the question *Where did X occur?* Clearly, however, the opposite analytical direction can be applied to the same dataset so that specific coordinates can be analyzed for non-spatial phenomena, instead answering the question *What occurred at X place?*

Geographic Information Systems are becoming increasingly utilized tools for analyses of cultural and linguistic data (see Derungs et al., 2018; Everett, 2013). Such studies typically target the location of whole languages or communities, or particular features representative of whole languages or communities, and explore, for example, their relationship to climate or topographical data. So far such mapping operates at a large-scale level, and the spatial analysis is therefore necessarily coarse-grained. The phenomena of study are statically represented entities or features, typically drawn from existing sources.

The methodological approach introduced here opens up the spatial inquiry into language and culture to a much finer level of granularity and precision, and to a very different type of data. Actual occurrences and instances of linguistic and cultural behavior – not just abstracted features – can be mapped and investigated with metric accuracy. This progression of the spatial analysis from features to occurrences parallels a general trend in the language sciences to generate and explore datasets of actual usage, as manifested in ever-growing language corpora and corpus-based research (cf. Newman et al., 2015). Our approach adds a geospatial dimension to this development. In this context, another difference worth emphasizing is the dynamic rather than static nature of our analytical environment. Explorable phenomena unfold in both time and space, allowing for a fine-grained analysis of the relationship between these two dimensions.

Our approach is of potential interest to any study or discipline concerned with the relationship between properties of observable behavior and geographical coordinates. The field of language documentation – with its agenda of generating audiovisual corpora and inherent interest in exhaustive records – comes to mind as one arena where it can be fruitfully employed. Audiovisually and spatially oriented research within anthropology, human geography, and political ecology are others. The technique is also suitable for exhaustive and efficient *in situ* tagging and commentary by researchers themselves, for example, during spatial surveys, where, say, categories or descriptions uttered by the surveyor can be linked to coordinates and examined geospatially.

We will conclude this discussion with an assessment of the ethical implications of our technique. Collection and integration of data of the kind described here raises ethical questions which have previously received limited attention. Ethical concerns and solutions related to recording and archiving of audiovisual materials have been abundantly discussed over the past 10–15 years, not least within the language documentation community (Thieberger & Musgrave, 2007). Similarly, there is a developing discourse about ethics and privacy issues in relation to spatial data, such as volunteered geographic citizen science data (Mooney et al., 2017). However, to our knowledge, the integration of dynamic audiovisual and geographic research data has not been addressed in the context of ethics and privacy concerns. Although our data integration does not add any recorded data type or format beyond the audiovisual and geospatial dimensions addressed in the above-mentioned literature, the combination of the two potently highlights the ability to register *who did or said what where, and when*. Apart from issues of privacy, participating individuals as well as researchers may, depending on the setting and context, find themselves entangled in spatio-cultural and spatio-political sensitivities of various kinds.

As in any ethics protocol, the issue will have to be addressed through rigid procedures of informed consent and data protection. But the complex combination of data does pose particular challenges, perhaps especially with regard to informed consent. In the context of language documentation, there is the additional aspect of long-term archiving, and the difficulties involved in foreseeing how documentary data will be used in the future (Thieberger & Musgrave, 2007), not to mention the expectations of making data openly accessible. In the absence of a specific ethics framework or precedent, we can at most offer a note of caution at this point, and encourage the relevant research communities to carefully consider the issue and share their thoughts and experiences. A useful data distinction to consider in this context is that between *volunteered*, *contributed* and *ambient* (or *passive*) geographic information (Goodchild, 2007; Harvey, 2013; Mooney et al., 2017; Stefanidis et al., 2013), which differ in the degree to which participants take an active part and are aware of the collection and use of their geospatial location. These distinctions are primarily employed to

characterize crowdsourcing, and one might want to consider adding a fourth category which signifies the profoundly participatory role community members play in the research process in many language documentation projects, for example, (cf. *Participatory GIS*, Abbot et al., 1998).

Indeed, our experiences from the indigenous communities in Southeast Asia involved in this study suggest that mobile recording technologies and the resulting data not only offer novel analytical environments but also open up new and exciting opportunities for community empowerment, and for involvement and autonomy in documentation and research.

Notes

1. <https://archive.mpi.nl/tla/elan>
2. http://www.mpi.nl/tools/elan/EAF_Annotation_Format_3.0_and_ELAN.pdf.
3. <https://www.w3.org/XML/>.
4. To revisit data collected with a traditional video camera and a dedicated GPS device, careful calibration, such as identifying shared locations, can be employed with similar results.
5. <https://www.thisisant.com/developer/resources/downloads/>.
6. <https://www.ogc.org/standards/kml>.

Acknowledgments

We are grateful to Assoc. Prof. Hajjah Rogayah A. Razak and Prof. Ulung Datuk Dr. Shamsul Amri Baharuddin, Universiti Kebangsaan Malaysia; the Ministry of Economic Affairs, Putrajaya; the Department of Orang Asli Development, Kuala Lumpur; the Lund University Humanities Lab; and the Penan, Jahai and Semelai communities.

Disclosure statement

The authors declare no competing interests.

Funding

This work was supported by the Bank of Sweden Tercentenary Foundation under Grants NHS14-1665:1 and IN17-0183:1; the Swedish Research Council under Grant HS-2016-25; and the Swiss National Science Foundation under Grant 162591.

Notes on contributors

Jens Larsson, MA, is Research Engineer at the Lund University Humanities Lab. He specializes in techniques of collection, management, and archiving of endangered languages data.

Niclas Burenhult is Assoc. Prof. of General Linguistics at Lund U. His research interests include the relationship between language, culture and cognition; spatial semantics; and language documentation.

Nicole Kruspe is a researcher in linguistics at Lund U. Her research focuses on the documentation and description of the Austroasiatic languages of Southeast Asia.

Ross S. Purves teaches and researches in the Department of Geography at the U. of Zurich. His interests include the use of language to better understand geography, and finding ways to incorporate the diversity of ways of thinking about space into geographical information systems.

Mikael Rothstein is Assoc. Prof. of Comparative Religion and Research Professor at the U. of Southern Denmark. His main research interests are religion among hunter-gatherers, religious emergence and innovation, and archaeology and religion.

Peter Sercombe is Senior Lecturer in Applied Linguistics at Newcastle U., UK. His research interests include sociolinguistics in relation to Austronesian languages, particularly effects of language contact.

References

- Abbot, J., Chambers, R., Dunn, C., Harris, T., de Merode, E., Porter, G., Townsend, J., & Weiner, D. (1998). Participatory GIS: Opportunity or oxymoron? *Participatory Learning and Action Notes*, 33, 27–34.
- Abernathy, D. (2016). *Using geodata and geolocation in the social sciences: Mapping our connected world*. SAGE.
- Ashmore, L. (2008). The role of digital video in language documentation. *Language Documentation and Description*, 5, 77–102. http://www.elpublishing.org/docs/1/05/ldd05_05.pdf
- Bates, C. (Ed.). (2015). *Video methods: Social science research in motion*. Routledge.
- Burenhult, N., & Purves, R. S. (2020). The spatial properties of forager motion categories: Evidence from Jahai. *Hunter-gatherer research*. in press.
- Cialone, C. (2019). *Placing spatial language and cognition in context through an investigation of Bininj Kunwok navigation talk* [Unpublished doctoral dissertation]. Australian National University.
- Derungs, C., Köhl, M., Weibel, R., & Bickel, B. (2018). Environmental factors drive language density more in food-producing than in hunter–gatherer populations. *Proceedings of the Royal Society B: Biological Sciences*, 285(1885), 20172851. <https://doi.org/10.1098/rspb.2017.2851>
- Everett, C. (2013). Evidence for direct geographic influences on linguistic sounds: The case of ejectives. *PLOS One*, 8(6), e65275. <https://doi.org/10.1371/journal.pone.0065275>
- Gawne, L., & Ring, H. (2016). Mapmaking for language documentation and description. *Language Documentation and Conservation*, 10, 188–242. <http://hdl.handle.net/10125/24692>
- Gianno, R. (2016). Landscape of ghosts: Semelai shamanism and a cosmological map. In K. M. Endicott (Ed.), *Malaysia's original people: Past, present and future of the Orang Asli* (pp. 203–225). National University of Singapore Press.
- Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4), 211–221. <https://doi.org/10.1007/s10708-007-9111-y>
- Harvey, F. (2013). To volunteer or to contribute locational information? Towards truth in labeling for crowdsourced geographic information. In D. Sui, S. Elwood, & M. Goodchild (Eds.), *Crowdsourcing geographic knowledge: Volunteered geographic information (VGI) in theory and practice* (pp. 31–42). Springer.
- Jewitt, C. (2012). An introduction to using video for research. *National Centre for Research Methods Working Paper 03/12*. London: Economic and Social Research Council.
- Margolis, E., & Pauwels, L. (Eds.). (2011). *The SAGE handbook of visual research methods*. SAGE.
- Mooney, P., Olteanu-Raimond, A.-M., Touya, G., Juul, N., Alvanides, S., & Kerle, N. (2017). Considerations of privacy, ethics, and legal issues in volunteered geographic information. In G. Foody, L. See, S. Fritz, P. Mooney, A.-M. Olteanu-Raimond, C. Costa Fonte, & V. Antoniou (Eds.), *Mapping and the citizen sensor* (pp. 119–135). Ubiquity Press.
- Newman, J., Baayen, H., & Rise, S. (Eds.). (2015). *Corpus-based studies in language use, language learning, and language documentation*. Brill.
- Rothstein, M. (2016). *Regnskovens religion. Forestillinger og ritualer blandt Borneos sidste jæger-samlere [The religion of the rainforest. Beliefs and rituals among Borneo's last hunter-gatherers]*. U Press.
- Sercombe, P., & Sellato, B. (Eds.). (2007). *Beyond the green myth. Borneo's hunter-gatherers in the twenty-first century*. NIAS Press.
- Sloetjes, H., & Seibert, O. (2016). Measuring by marking: The multimedia annotation tool ELAN. In A. Spink, G. Riedel, L. Zhou, L. Teekens, R. Albatal, & C. Gurrin (Eds.), *Proceedings of measuring behavior 2016* (pp. 492–495). Dublin City University.
- Spencer, J., Frizzelle, B. G., Page, P. H., & Vogler, J. B. (2003). *Global positioning system: A field guide for the social sciences*. Blackwell.
- Spencer, S. (2010). *Visual research methods in the social sciences: Awakening visions*. Routledge.
- Stefanidis, A., Crooks, A., & Radzikowski, J. (2013). Harvesting ambient geospatial information from social media feeds. *GeoJournal*, 78(2), 319–338. <https://doi.org/10.1007/s10708-011-9438-2>
- Thieberger, N., & Musgrave, S. (2007). Documentary linguistics and ethical issues. *Language Documentation and Description*, 4, 26–37. http://www.elpublishing.org/sites/default/files/attachments/ldd04_03.pdf
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In N. Calzolari, K. Choukri, A. Gangemi, B. Maegaard, J. Mariani, J. Odiijk, & D. Tapias, (Eds) *Proceedings of the Fifth International Conference on Language Resources and Evaluation* (pp. 1556–1559). Genoa: European Language Resources Association.